Discrete Mathematics 11 (1975) 125–132.

SHORTEST STRING CONTAINING ALL PERMUTATIONS*

P.J. KOUTAS and T.C. HU

Mathematics Research Center, The University of Wisconsin, Madison, Wisc. 53706, USA

Received 23 January 1973
Revised 17 May 1974

In this paper, we consider the problem of constructing a shortest string of $\{1, 2, ..., n\}$ containing all permutations on n elements as subsequences. For example, the string 1 2 1 3 1 2 1 contains the 6 (= 3!) permutations of $\{1, 2, 3\}$ and no string with less than 7 digits contains all the six permutations. Note that a given permutation, such as 1 2 3, does not have to be consecutive but must be from left to right in the string.

We shall first give a rule for constructing a string of $\{1, 2, ..., n\}$ of infinite length and then show that the leftmost $n^2 - 2n + 4$ digits of the string contain all the $n!$ permutations (for $n \geq 3$). We conjecture that the number of digits $f(n) = n^2 - 2n + 4$ (for $n \geq 3$) is the minimum.

Then we study a new function $F(n, k)$ which is the minimum number of digits required for a string of n digits to contain all permutations of i digits, $i \leq k$. We conjecture that $F(n, k) = k(n-1)$ for $4 \leq k \leq n-1$.

In this paper, we consider the problem of constructing a shortest string of $\{1, 2, ..., n\}$ containing all permutations on n elements as subsequences. For example, the string 1 2 1 3 1 2 1 contains the 6 (= 3!) permutations of $\{1, 2, 3\}$ and no string with less than 7 digits contains all the six permutations. Note that a given permutation, such as 1 2 3, does not have to be consecutive but must be from left to right in the string.

For $n = 4$, a shortest string has 12 digits and for $n = 5$, Newey [2] claims that a shortest string must have 19 digits. The problem of finding a shortest string of $\{1, 2, ..., n\}$ was first proposed by R.M. Karp. It is also listed as an open problem in [1].

We shall first give a rule for constructing a string of $\{1, 2, .., n\}$ of finite length and then show that the leftmost $n^2 - 2n + 4$ digits of the string contain all the $n!$ permutations (for $n \geq 3$). We conjecture that

* Sponsored by the United States Army under Contract No. DA-31-124-ARO-D-462 and National Science Foundation Grant GJ-28339A1.

the number of digits $f(n) = n^2 - 2n + 4$ (for $n \geq 3$) is the minimum.

Consider the mapping of n digits

$$\begin{pmatrix} 1, 2, 3, ..., n-1, n \\ 1, n, 2, ..., n-2, n-1 \end{pmatrix}.$$

Note that the mapping contains two cycles, a cycle of length one and a cycle of length $n-1$, which contains the digits $2, 3, ..., n$. If we start with the sequence $1, 2, 3, ..., n$, and apply the mapping k times, we have $1, n-k+1, n-k+2, ..., n, 2, ..., n-k$. We shall denote the sequence obtained after k mappings as $D_k(n)$. Thus, we have:

$$D_0(n) = 1, 2, 3, ..., n;$$
$$D_k(n) = 1, n-k+1, n-k+2, ..., n, 2, ..., n-k;$$
$$D_{n-1}(n) = 1, 2, 3, ..., n.$$

The subsequence containing the leftmost m digits of the sequence $D_k(n)$ is denoted by $D_k(m \mid n)$. It is clear that $D_p(m \mid n) = D_q(m \mid n)$ since $D_p(n) = D_q(n)$ for $p \equiv q \pmod{n-1}$.

Now we can give a simple rule for constructing an infinite string of $\{1, 2, ..., n\}$. The infinite string contains the following sequences from left to right:

(1) $D_0(n), D_0(n-1 \mid n), D_1(n-1 \mid n), D_2(n-1 \mid n), ..., D_{n-1}(n-1 \mid n), ...$

If we omit the sequence $D_0(n)$ from the infinite string, the infinite string repeats the subsequence $D_0(n-1 \mid n), D_1(n-1 \mid n), ..., D_{n-2}(n-1 \mid n)$ infinitely.

Applying this rule, we have:

1234, 123, 142, 134, 123, ...	for $n = 4$,
12345, 1234, 1523, 1452, 1345, 1234, ...	for $n = 5$,
123456, 12345, 16234, 15623, 14562, 13456, 12345, ...	for $n = 6$.

In the infinite string, we say that the distance from the pth position to the qth position is $q-p$ for $1 \leq p < q$. If we fix our attention on a particular digit i in the pth position in the infinite string, we shall define the distance from i to another digit j where j is on the right of i. Note that there is an infinite number of digits j (on the right of i) occupying the positions $q_1, q_2, ..., q_t, ...$. The distance from a digit i in the pth position to a digit j is defined to be $\min_t(q_t - p)$. That is to say, we consider only the first occurrence of the digit j after the pth position.

Lemma 1. *In the infinite string defined by* (1), *the distance from a digit* 1 *to a digit* j ($j \neq 1$) *is at most* n. *The distance from a digit* i ($i \neq 1$) *to a digit* j ($j \neq i$) *is at most* $n-1$.

Proof. Consider the infinite string

$$D_0(n), D_0(n-1 \mid n), D_1(n-1 \mid n), D_2(n-1 \mid n), \ldots, D_{n-2}(n-1 \mid n), D_{n-1}(n-1 \mid n), \ldots .$$

If the digit 1 is in the first position of $D_0(n)$, we see that the next $n-1$ positions of $D_0(n)$ contain $n-1$ distinct digits, and the distance from 1 to j is at most $n-1$. If the digit 1 is the first digit of $D_k(n-1 \mid n)$, then the next $n-2$ positions contain $n-2$ distinct digits different from 1, and the digit not in $D_k(n-1 \mid n)$ occupies the second position of $D_{k+1}(n-1 \mid n)$. This follows from the definition of the mapping. Thus the distance is at most $(n-2)+2=n$.

Now consider a digit i ($i \neq 1$) occupying the pth position. The next $n-1$ positions after this i always contain $n-1$ distinct digits different from i. Hence the distance is at most $n-1$. This completes the proof.

Theorem 1. *The first* $f(n)$ *digits of the infinite string of* $\{1, 2, \ldots, n\}$ *defined by* (1) *contain all* $n!$ *permutations of* n, *where* $f(n) = n^2 - 2n + 4$ ($n \geqslant 3$).

Proof. For a generic permutation, let us consider the position of the digit 1. The digit 1 is

(i) the first digit of the permutation,

(ii) the last digit of the permutation, or

(iii) the tth digit of the permutation.

In each of the three cases, we shall count the number of digits in the string which will contain the given permutation.

(i) The digit "1" is the first digit of the permutation.

Since the first digit of the string is 1 and the next $n-1$ digits are distinct and different from 1, we need at most $1+(n-1)=n$ digits in the string for the first two digits of the permutation. For the remaining $n-2$ digits, each digit j is a distance $n-1$ or less from its preceding digit i ($i \neq 1$) (see Lemma 1). Thus we need at most $1+(n-1)+(n-2)(n-1) = n^2-2n+2$ digits in the string.

(ii) The digit "1" is the last digit of the permutation.

Consider the subsequences in the infinite string

$$D_0(n), D(n-1\,|\,n), D_1(n-1\,|\,n), D_2(n-1\,|\,n), \ldots .$$

The first digit of the permutation must be contained in $D_0(n)$.

For the next $n-2$ digits, each digit is at most $n-1$ from its preceding digit, we need a total of $n+(n-2)(n-1)=n^2-2n+2$ digits. Now, there are n^2-2n+2 digits in the first $n-1$ subsequences of the finite string; namely,

$$D_0(n), D_0(n-1\,|\,n), D_1(n-1\,|\,n), \ldots, D_{n-3}(n-1\,|\,n).$$

But the first digit of $D_{n-2}(n-1\,|\,n)$ is "1". Hence we need at most $(n^2-2n+2)+1$ digits in the string.

(iii) The digit "1" is in the tth digit of the permutation.

Just as in case (ii) we need $n+(t-2)(n-1)+1$ digits for the first t digits of the permutation. From Lemma 1, the $(t+1)$th digit of the permutation is at a distance $\leqslant n$ from 1. For remaining $[n-(t+1)]$ digits, each is at distance $\leqslant n-1$ from its preceding digit. Thus, we have

$$n+(t-2)(n-1)+1+n+(n-t-1)(n-1)=n^2-2n+4$$

digits. Therefore, the leftmost $f(n)$ digits of the infinite string contain all $n!$ permutations of n, where $f(n)=n^2-2n+4$, $n\geqslant 3$. Note that only for $n\geqslant 3$, the digit 1 can occupy a position which is not the first or the last position in the permutation.

Let $F(n, 1)$ denote the minimum number of digits in a string of n digits which contains all single digits. In this notation, $F(n, 1)=n$, and the string starts with n distinct digits. If we add more digits to the right of the n distinct digits, then, eventually, the string will contain all permutations of two digits. If we keep the addition to a minimum, then we denote the total number of digits by $F(n, 2)$. It is easy to see that $F(n, 2)=2n-1$. If we keep adding digits successively, so that the string contains all permutations of k digits, $k=3, 4, \ldots$, and each time keeping the addition to a minimum, eventually, we will have a total of $F(n, n)$ digits and the strings contain all permutations of i digits for all $i\leqslant n$. In general, $F(n, n)\neq f(n)$. Because in a shortest string of $f(n)$ digits, we are only concerned about permutations of n digits, not concerned with permutations of i digits for $i<n$. Furthermore, we do not

have to build the string successively from left to right as we did for $F(n, n)$. We shall prove the following

(2) $F(n, 1) = n$,

(3) $F(n, 2) = 2n - 1$,

(4) $F(n, 3) = 3n - 2$,

(5) $F(n, 4) = 4n - 4$,

(6) $F(n, 5) = 5n - 5$,

and we conjecture that $F(n, k) = k(n - 1)$ for $4 \leqslant k \leqslant n - 1$.

In order to prove the above, we construct a game as follows. There are two players S and P. The player S supplies the digits in the string and he wants to achieve the minimum number of digits. The player P picks the permutation and he tries to make S use the maximum number of digits.

Lemma 2. $F(n, 1) = n$.

Proof. The player P looks at the leftmost n digits in the string, if they are all distinct, then P would pick the nth digit in the string as his first digit, call it x_n. If the leftmost n digits are not distinct, player P can continue to look to the right and keep track of which digits have already appeared in the string. Eventually, there will be only one digit which has no yet appeared. Then player P would pick that digit as his first digit x_n. This proves the lemma.

Since x_n occupies the nth position in the string, we shall call the first n positions the first interval, and the first $n-1$ positions the first open interval, and the position occupied by x_n the first end position.

Lemma 3. $F(n, 2) = 2n - 1$.

Proof. Since the string must contain n distinct digits $x_1, x_2, \ldots, x_n$ in the first interval, the player P can pick x_n as his first digit and then use the same strategy as in Lemma 1, thus S has to supply $n-1$ distinct digits to the right of x_n. Call the right most digits x_{n-1}. Then we have $F(n, 2) \geqslant n + (n-1) = 2n-1$. It is easy to see that $2n-1$ is also sufficient. Thus

$F(n, 2) = 2n-1$. Using the terminology introduced above, we have $n-1$ distinct digits in the second interval, the digit $[x_n]$ in the first end position is different from the digit in the second end position, and $[x_n]$ is different from all the digits in the two adjacent open intervals (i.e., the first and the second open intervals).

Lemma 4. $F(n, 3) = 3n-2$.

Proof. A string satisfying Lemmas 2 and 3 can be written symbolically as $x_1, \ldots, [x_n], \ldots, [x_{n-1}]$, where we have used brackets around the digits in the first and the second end positions. The player P can pick $[x_n]$ and $[x_{n-1}]$ as the first and the second digits, and then any digit from $\{x_1, x_2, \ldots, x_{n-2}\}$ as the third digit. This shows that we need at least $n-2$ digits to the right of $[x_{n-1}]$. But we also need x_n to the right of $[x_{n-1}]$. This can be seen as follows. Let x_r and x_s be any two digits in the first open interval where $x_r < x_s$ (we write $x_r < x_s$ if x_r is on the left of x_s). Then the player P can pick x_s as the first digit, x_r as the second digit (note that x_r picked is in the second open interval) and then x_n as the third digit. Thus x_n must also be added to the right of $[x_{n-1}]$. Therefore $F(n, 3) = (2n-1) + (n-2) + 1 = 3n-2$. Note that a digit in an end position is always different from any digit in the two adjacent intervals, and all digits in an open interval are distinct.

Lemma 5. $F(n, 4) = 4n-4$.

Proof. A string satisfying Lemmas 2, 3 and 4 can be written symbolically as $x_1, x_2, \ldots, [x_n], \ldots, [x_{n-1}], \ldots, x_p, [x_{n-2}]$, where x_p is the rightmost digit in the third open interval. We shall consider two cases.

Case 1: $[x_{n-2}] = x_n$. In this case, player P can pick x_n, x_{n-1} and x_p as the first three digits. Then the player S needs to put $(n-3)$ additional digits to the right of $[x_{n-2}]$. In addition to the $(n-3)$ digits, player S needs also to put x_p and x_{n-1} to the right of $[x_{n-2}]$. This can be seen as follows.

Let x_r, x_s be two digits in the first open interval, where $x_r < x_s$ and $x_r, x_s \neq x_n, x_{n-1}, x_p$. Then the player P can pick x_s, x_r and $x_n = [x_{n-2}]$ (which occupies the third end position). Then the player P can pick either x_p or x_{n-1} as the fourth digit. This shows that both x_p and x_{n-1} would be needed if $[x_{n-2}] = x_n$. In this case, the total number of digits would be

$$(3n-2) + (n-3) + 2 = 4n-3.$$

Case 2: $[x_{n-2}] \neq x_n$. Obviously, $(n-3)$ new digits (excluding x_n, x_{n-1}, x_{n-2}) have to be on the right of $[x_{n-2}]$.

Let x_r, x_s be two digits in the first open interval,

$$x_r < x_s \quad \text{and} \quad x_r, x_s \neq x_n, x_{n-1}.$$

Then the player P can pick x_s, x_r, x_n (in the third open interval) and then x_{n-1}. In this case, we need a total of

$$(3n-2)+(n-3)+1 = 4n-4$$

digits. Now we shall exhibit a string of $3n-3$ digits, which symbolically could be as follows (these are the leftmost digits of the $4n-4$ digits which achieves $F(n, 4)$):

$$\underbrace{x_1, x_2, \ldots, x_{n-1}[x_n]}_{n \text{ digits}}, \underbrace{\ldots, x_{n-2}[x_{n-1}]}_{n-1 \text{ digits}}, \underbrace{\ldots, x_n[x_{n-2}]}_{n-1 \text{ digits}}.$$

To show that this string does contain all portions of 4 digits, we have to show that x_n is not needed on the right of $[x_{n-2}]$. If x_n is needed to the right of $[x_{n-2}]$, then the third digit that the player P picked must be $[x_{n-2}]$ (since x_n is immediately to its left). Because all digits are distinct in the third open interval, the second digit must be $[x_{n-1}]$. Then the first digit must be $[x_n]$. So it is sufficient to have $[x_n]$, $[x_{n-1}]$, $[x_{n-2}]$ as the first three digits and then pick any $x \in \{x_1, x_2, \ldots, x_{n-3}, x_{n-1}\}$ to the right of $[x_{n-2}]$.

On the other hand, if we do not pick $[x_n]$ as the first digit, the first three digits need a string of $3n-2$ or less, and the third digit is to the left of $x_n[x_{n-2}]$. Then the fourth digit could be any one of the n digits. Thus $F(n, 4) = 4n-4$.

Lemma 6. $F(n, 5) = 5n-5$.

A string satisfying Lemmas 2, 3, 4 and 5 can be written symbolically as $x_1, \ldots, x_{n-1}, [x_n], \ldots, x_{n-2}, [x_{n-1}], \ldots, x_n, [x_{n-2}], \ldots, [x_{n-3}]$.

Case 1: $[x_{n-3}] \neq x_{n-1}$. The player P may pick $[x_n]$, $[x_{n-1}]$, $[x_{n-2}]$, $[x_{n-3}]$ as his first four digits and pick any digit from the remaining $n-4$ digits. Thus, we need at least these $n-4$ digits to the right of $[x_{n-3}]$. In addition, we also need x_n, x_{n-1} and x_{n-2} to the right of $[x_{n-3}]$.

Let x_r, x_s, x_p (and $x_r, x_s, x_p \neq x_n, x_{n-1}$) be three digits in the second interval, where $x_r < x_p, x_s < x_p$ in the second interval and $x_r < x_s$ in the third interval. Then the player P can pick x_{n-1}, x_p, x_s, x_r (in the fourth interval) and then x_n. This shows that x_n is needed to the right of $[x_{n-3}]$. Using similar reasoning, we can also show that x_{n-1} and x_{n-2} are both needed to the right of $[x_{n-3}]$.

Case 2: $[x_{n-3}] = x_{n-1}$. This case is similar to Case 1. In both cases, we have $F(n, 5) = 5n-5$.

Lemma 7. *If $F(n, k) = m$, then $F(n+1, k) \leqslant m+k$.*

Proof. When a new digit is added to the string, we can add k of the new digit, each in the beginning of each open interval. Then it is sufficient to have all permutations of k digits out of the $(n+1)$ digits.

At this point, it is very tempting to conjecture that

$$(7) \qquad F(n, k) = k(n-1) \quad \text{for } 4 \leqslant k \leqslant n-1.$$

It is clear that $F(n, n) \geqslant f(n)$. We would like to show that $F(n, n) > f(n)$. For example, $F(7, 4) = 24$ and the string is

$$1,2,3,4,5,6,[7],1,2,3,4,5,[6],1,2,3,4,7,[5],1,6,2,3,[4].$$

By exhaustive search, it seems impossible to add another 15 digits to achieve $F(n, n)$. This shows that $F(n, n) > f(n) = 24 + 15 = 39 = n^2 - 2n + 4$ (for $n = 7$).

References

[1] V. Chvátal, D.A. Klarner and D.E. Knuth, Selected combinatorial research problems, Computer Science Department Report, CS-72-292, Stanford University, Stanford, Calif. (June 1972).

[2] M. Newey, Notes on a problem involving permutations as subsequences, Computer Science Department Report, CS-73-340, Stanford University, Stanford, Calif. (March 1973).